PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

# INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : G10L 5/00

A1

(11) International Publication Number: WO 86/ 03872

(43) International Publication Date: 3 July 1986 (03.07.86)

(21) International Application Number: PCT/US85/02448

(22) International Filing Date: 11 December 1985 (11.12.85)

(31) Priority Application Numbers: 684,382; 798,174

(32) Priority Dates: 20 December 1984 (20.12.84); 14 November 1985 (14.11.85)

(33) Priority Country: US

(71) Applicant: GTE LABORATORIES INCORPORATED [US/US]; 100 West 10th Street, Wilmington, DE 19800 (US).

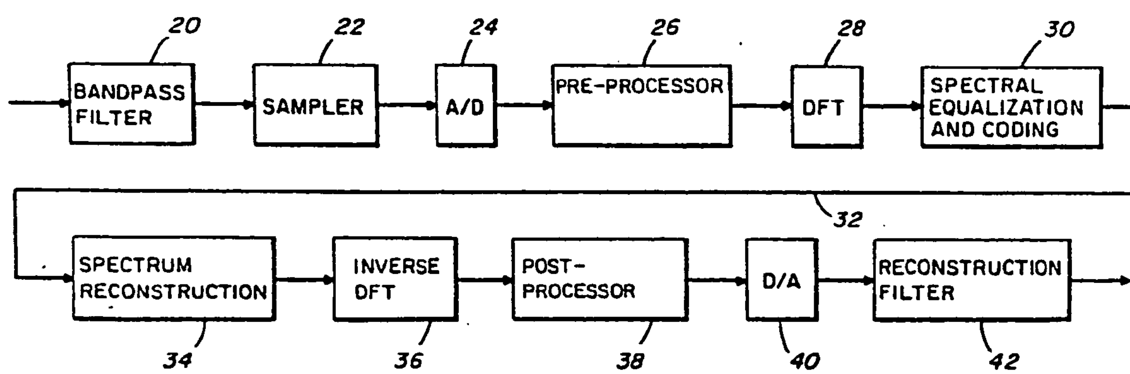
(72) Inventors: ZIBMAN, Israel, Bernard ; 423 Ward Street, Newton, MA 02159 (US). MAZOR, Baruch ; 73 John Street, Newton Centre, MA 02159 (US). VEENEMAN, Dale, E. ; 269 Cordaville Road, Southborough, MA 01722 (US).

(74) Agent: KEAY, David, M.; GTE Service Corporation, 100 First Avenue, Waltham, MA 02254 (US).

(81) Designated States: BE (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent).

**Published**
*With international search report.*

(54) Title: ADAPTIVE METHOD AND APPARATUS FOR CODING SPEECH




(57) Abstract

In a speech encoder, a Fourier transform (28) of the speech is provided. The Fourier transform is equalized (30) by normalizing the spectrum coefficients to a curve which approximates the shape of the spectrum. Both the curve and the equalized spectrum are encoded. In one system, scale factors (45) are generated and encoded for each of a plurality of subbands of a Fourier transform spectrum of speech. Based on those scale factors, the spectrum is equalized (46). Coefficients of a limited number of subbands (48) determined by the scale factors are encoded (50). The number of bits used to encode each coefficient of each transmitted subband is determined by the scale factor for each subband. At the receiver, coefficients of subbands which are not transmitted are approximated by means of a list replication technique (54).

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AT | Austria |
| AU | Australia |
| BB | Barbados |
| BE | Belgium |
| BG | Bulgaria |
| BR | Brazil |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CM | Cameroon |
| DE | Germany, Federal Republic of |
| DK | Denmark |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| HU | Hungary |
| IT | Italy |
| JP | Japan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LU | Luxembourg |
| MC | Monaco |
| MG | Madagascar |
| ML | Mali |
| MR | Mauritania |
| MW | Malawi |
| NL | Netherlands |
| NO | Norway |
| RO | Romania |
| SD | Sudan |
| SE | Sweden |
| SN | Senegal |
| SU | Soviet Union |
| TD | Chad |
| TG | Togo |
| US | United States of America |

# ADAPTIVE METHOD AND APPARATUS FOR CODING SPEECH

The present invention relates to digital coding of speech signals for telecomunications and has particular application to systems having a transmission rate of about 16,000 bits per second or less.

Conventional analog telephone systems are being replaced by digital systems. In digital systems, the analog signals are sampled at a rate of about twice the bandwidth of the analog signals or about eight kilohertz, and the samples are then encoded. In a simple pulse code modulation system (PCM), each sample is quantized as one of a discrete set of prechosen values and encoded as a digital word which is then transmitted over the telephone lines. With eight bit digital words, for example, the analog sample is quantized to $2^8$ or 256 levels, each of which is designated by a different eight bit word. Using nonlinear quantization, excellent quality speech can be obtained with only seven bits per sample; but since a seven bit word is still required for each sample, transmission bit rates of 56 kilobits per second are necessary.

Efforts have been made to reduce the bit rates required to encode the speech and obtain a clear decoded speech signal at the receiving end of the system. The linear predictive coding (LPC) technique is based on the recognition that speech production involves excitation and a filtering process. The excitation is determined by the vocal cord vibration for voiced speech and by turbulence for unvoiced speech, and that actuating signal is then modified by the filtering process of vocal resonance chambers, including the mouth and nasal passages. For a particular group of samples, a digital filter which simulates the formant effects of the resonance chambers can be defined and the definition can be encoded. A

residual signal which approximates the excitation can then be obtained by passing the speech signal through an inverse formant filter, and the residual signal can be encoded. Because sufficient information is contained in the lower-frequency portion of the residual spectrum, it is possible to encode only the low frequency baseband and still obtain reasonably clear speech. At the receiver, a definition of the formant filter and the residual baseband are decoded. The baseband is repeated to complete the spectrum of the residual signal. By applying the decoded filter to the repeated baseband signal, the initial speech can be reconstructed.

A major problem of the LPC approach is in defining the formant filter which must be redefined with each window of samples. A complex encoder and a complex decoder are required to obtain transmission rates as low as 16,000 bits per second. Another problem with such systems is that they do not always provide a satisfactory reconstruction of certain formants such as that resulting, for example, from nasal resonance.

In accordance with the present invention, speech is encoded by first performing a transform of a window of speech. Preferably the transform is the Fourier transform. The discrete transform spectrum is normalized by defining at least one curve approximating the magnitude of the discrete spectrum, digitally encoding the defined curve and redefining the discrete spectrum relative to the defined curve to provide a normalized spectrum. More specifically, the defined curve is the approximate envelope of the discrete spectrum. Preferably, the discrete spectrum is normalized by determining the maximum magnitude of the spectrum within each of a plurality of regions of the spectrum, digitally encoding the maximum magnitude of each region and redefining the spectrum by scaling each coefficient of the spectrum in each region to

the maximum magnitude of that region. At least a portion of the normalized spectrum is then encoded.

In one system, the approximate envelope of the transform spectrum in each of a plurality of subbands of coefficients is defined and each envelope definition is encoded for transmission. Each spectrum coefficient is then scaled relative to the defined envelope of the respective subband, and each scaled coefficient is encoded in a number of bits which is determined by the defined envelope of its subband.

Zero bits may be allotted to a number of less significant subbands as indicated by the defined envelopes; and varying numbers of bits may be used for each encoded coefficient depending on the magnitude of the defined envelope for the respective subband. Thus, the subbands which are transmitted and the resolution with which the transmitted subbands are encoded are determined adaptively for each sample window based on the defined envelopes of the subbands.

At the receiver, the subbands which are transmitted are replicated to define coefficients of frequencies which are not transmitted. A list replication procedure is followed by which an nth coefficient which is transmitted is replicated as an nth coefficient which is not transmitted. After replication the speech signal can be recreated by using the transmitted envelope definitions to inverse scale the coefficients of the respective subbands and by performing an inverse transform.

In another system the spectrum is normalized first with respect to only a few regions and subsequently with respect to a greater number of subregions. The maximum magnitude in each of the regions and in each of the subregions is encoded. The maximums are logarithmically encoded and only a baseband of the normalized spectrum is encoded.

The foregoing and other objects, features, and advantages of the invention will be apparent from the following more particular description of a preferred embodiment of the invention, as illustrated in the accompanying drawings in which like reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

Fig. 1 is a block diagram illustration of an encoder and a decoder embodying the present invention;

Figure 2 is a block diagram of a speech encoder and corresponding decoder of a preferred implementation of the system of Figure 1.

Figure 3 is an example of a magnitude spectrum of the Fourier transform of a window of speech illustrating principles of the system of Figure 2.

Figure 4 is an example spectrum normalized from that of Figure 3 based on principles of the present invention.

Figure 5 schematically illustrates a quantizer for complex values of the normalized spectrum.

Figure 6 is an example illustration of coefficient groups which are transmitted and illustrates the replication technique of the system of Figure 2.

Figure 7 is an example of a magnitude spectrum of a window of speech illustrating principles of another system embodying the present invention.

Figure 8 is an example spectrum normalized from the spectrum of Fig. 7 using four formant regions;

Figure 9 is an example spectrum normalized from that of Fig. 8 in subbands;

Figure 10 schematically illustrates a quantizer for complex values of the normalized spectrum;

Figure 11 is a block diagram illustration of the spectral equalization encoding circuit of Fig. 1 in the alternative embodiment.

A block diagram of the system is shown in Fig. 1. Speech is filtered with a telephone bandpass filter 20 which prevents aliasing when the signal is sampled 8,000 times per second in sampling circuit 22. The analog samples are digitally encoded in an analog to digital encoder 24 and are preprocessed at 26 prior to being applied to a discrete Fourier transform unit 28.

The output of the Fourier transform circuit 28 is a sequence of coefficients which indicate the magnitude and phase of the Fourier transform spectrum at each of 97 frequencies spaced 41.667 hertz apart. The magnitude spectrum of the Fourier transform output is illustrated as a continuous function in Fig. 3 but it is recognized that the transform circuit 28 would actually provide only 97 incremental outputs.

In accordance with the present invention, the Fourier transform spectrum of the full speech within a selected window is equalized and encoded in circuit 30 in a manner which will be discussed below. The resultant digital signal can be transmitted at 16,000 bits per second over a line 32 to a receiver. At the receiver the full spectrum of Fig. 3 is reconstructed in circuit 34. The inverse Fourier transform is performed in circuit 36 and applied through a post-processor 38 corresponding to the pre-processor 26. That signal is then converted to analog form in digital to analog converter 40. Final filtering in filter 42 provides clear speech to the listener.

In a preferred system, a pipelined multiprocessor architecture is employed. One microcomputer is dedicated to the analog to digital conversion with preemphasis filtering, one is dedicated to the forward Fourier transform and a third is dedicated to the spectral equalization and coding. Similarly, in the receiver, one microcomputer is dedicated to spectrum reconstruction, another to inverse Fourier transform and a third to digital to analog conversion with deemphasis filtering.

The spectral equalization and encoding technique of the present invention is based on the recognition that the Fourier transform of the total signal includes a relatively flat spectrum of the pitch illustrated in Fig. 4 shaped by formant signals. In the present system, the signal of Fig. 4 is obtained by normalizing the spectrum of Fig. 3 to at least one curve which itself can be encoded separate from the residual spectrum of Fig. 4.

One implementation of the coding system of Figure 1 is shown in Figure 2. Prior to compression, the analog speech signal is low pass filtered in filter 20 at 3.4 kilohertz, sampled in sampler 22 at a rate of 8 kilohertz, and digitized using a 12 bit linear analog to digital converter 24. It will be recognized that the input to the encoder may already be in digital form and may require conversion to the code which can be accepted by the encoder. The digitized speech signal, in frames of N samples, is first scaled up in a scaler 26 to maximize its dynamic range in each frame. The scaled input samples are then Fourier transformed in a fast Fourier transform device 28 to obtain a corresponding discrete spectrum represented by (N/2)+ 1 complex frequency coefficients.

In a specific implementation, the input frame size equals 180 samples and corresponds to a frame every 22.5 milliseconds. However, the discrete Fourier transform is performed on 192 samples, including 12 samples overlapped with the previous frame, preceded by trapezoidal windowing with a 12 point slope at each end. The resulting output of the FFT includes 97 complex frequency coefficients spaced 41.667 Hertz apart.

An example magnitude spectrum of a Fourier transform output from FFT 28 is illustrated in Figure 2. Although illustrated as a continuous function, it is recognized that the transform circuit 28 actually provides only 97 incremental complex outputs.

The magnitude spectrum of the Fourier transform output is equalized and encoded. To that end, the spectrum is partitioned into contiguous subbands and a spectral envelope estimate is based on a piecewise approximation of those subbands at 44. In a specific implementation, the spectrum is divided into twenty subbands, each including four complex coefficients. Frequencies above 3291.67 Hertz are not encoded and are set to zero at the receiver. To equalize the spectrum, the spectral envelope of each subband is assumed constant and is defined by the peak magnitude in each subband as illustrated by the horizontal lines in Figure 3. Each magnitude, or more correctly the inverse thereof, can be treated as a scale factor for its respective subband. Each scale factor is quantized in a quantizer 45 to four bits.

By then multiplying at 46 the magnitude of each coefficient of the spectrum by the scale factor associated with that coefficient, the flattened residual spectrum of Figure 4 is obtained. This flattening of the spectrum is equivalent to inverse filtering the signal based on the piecewise-constant estimate of the spectral envelope.

Only selected subbands of the flattened spectrum of Figure 4 are quantized and transmitted. Selection at 48 of subbands to be transmitted is based on the scale factor of the subbands. In a specific implementation, the 12 subbands having the smallest scale factors, that is the largest energy, are encoded and transmitted. For the eight lower energy subbands only the scale factors are transmitted.

A nonuniform bit allocation is used for the complex coefficients which are transmitted. Three separate two dimensional quantizers 50 are used for the transmitted 12 subbands. The sixteen complex coefficients of the four subbands having the smallest scale factors are quantized to seven bits each. The coefficients of the four subbands

having the next smallest scale factors are quantized to six bits each, and the coefficients of the remaining four of the transmitted subgroups are quantized to four bits each. In effect, the coefficients of the eight subbands which are not transmitted are quantized to zero bits.

Each of the two dimensional quantizers is designed using an approach presented by Linde, et al., "An Algorithm for Vector Quantizer Design," IEEE Trans on Commun, Vol COM-28, pp. 84-95, Jan 1980. The result for the seven bit quantizer is shown in Figure 5. The two dimensions of the quantizer are the real and imaginary components of each complex coefficient. Each cluster has a seven bit representation to which each complex point in the cluster is quantized. Actual quantization may be by table look-up in a read only memory.

The bit allocation for a single frame may be summarized as follows:

| | | |
|---|---|---|
| Scale factors | 20 x 4 bits each = | 80 bits |
| | 16 x 7 bits = | 112 bits |
| | 16 x 6 bits = | 96 bits |
| | 16 x 4 bits = | 64 bits |
| | Time scaling = | 4 bits |
| | Synchronization = | 4 bits |
| | TOTAL | 360 bits |

At the receiver, the transmitted 12 groups of coefficients are applied to corresponding seven bit, six bit and four bit inverse quantizers at 52. The frequency subbands to which the resulting coefficients correspond are determined by the scale factors which are transmitted in sequence for all subbands. Thus, the coefficients from the seven bit inverse quantizer are placed in the subbands which the scale factors indicate to be of the greatest magnitude.

The coefficients of the eight subbands which are not transmitted are approximated by replication of transmitted subbands at 54. To that end, a list replication approach is utilized. This approach is illustrated by Figure 6. In Figure 6, the coefficients for each subband are illustrated by a single vector. The transmitted subbands are indicated as T1, T2, T3, . . .Tn, . . . and the subbands which must be produced by replication in the receiver are indicated as R1, R2, R3, . . . Rn, . . . In accordance with the replication technique of the present system, the coefficients of the subband Tn are used both for Tn and for Rn. Thus, the scaled coefficients for subband T1 are repeated at subband R1, those of subband T2 are repeated at R2, and those at subband T3 are repeated at R3. The rationale for this list replication technique is that subbands are themselves usually grouped in blocks of transmitted subbands and blocks of nontransmitted subbands. Thus, large blocks of coefficients are typically repeated using this approach and speech harmonics are maintained in the replication process.

Once the equalized spectrum of Figure 4 is recreated by replication of subbands, a reproduction of the spectrum of Figure 3 can be generated at 56 by applying the scale factors to the equalized spectrum. From that Fourier transform reproduction of the original Fourier transform, the speech can be obtained through an inverse FFT 36, an inverse scaler 38, a digital to analog converter 40 and a reconstruction filter 42.

A distinct advantage of the present system is that the coder is not based on an assumed fixed low pass spectrum model which is speech specific. Voice-band data and signaling take the form of sine waves of some bandwidth which may occur at any frequency. Where only a lower or an upper baseband of coefficients is transmitted, voice-band data can be lost. With the present system, the

subbands in which digital information is transmitted are naturally selected because of their higher energy.

Another attractive feature of the coding system is its embedded data-rate codes capability. Embedded coding, important as a method of congestion control in telephone applications, allows the data to leave the encoder at a constant bit rate, yet be received at the decoder at a lower bit rate as some bits are discarded enroute. Embedded coding implies a packet or block of bits within which there is a hierarchy of subblocks. Least crucial subblocks can be discarded first as the channel gets overloaded. This hierarchical concept is a natural one in the present system where the partial-band information, described by a set of frequency coefficients, is ordered in a decreasing significance and the missing coefficients can always be approximated from the received ones. The more coefficients in the set, the higher is the rate and the better is the quality. However, speech quality degrades very gracefully with modest drops in the rate. The implementation of an embedded coding system in conjunction with this approach is therefore fairly simple and very attractive.

The coding technique described above provides for excellent speech coding and reproduction at 16 kilobits per second. Excellent results as low as 8.0 kilobits per second can be obtained by using this technique in conjunction with a frequency scaling technique known as time domain harmonic scaling and described by D. Malah, "Time Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, pp. 121-133, Apr. 1979. In that approach, prior to performing the fast Fourier transform, speech at twice the rate of the original speech but at the original pitch is generated by combining adjacent pitch cycles. The frequency scaled

speech can then be fast Fourier transformed in the technique described above.

Although each of the steps of residual extraction, subband selection, and quantizing and the steps of inverse quantizing, replication and envelope excitation are shown as individual elements of the system, it will be recognized that they can be merged in an actual system. For example, the residual spectrum for subbands which are not transmitted need not be obtained. The system can be implemented using a combination of software and hardware.

In another coding system, the shape of the spectrum is determined by a two-step process. This process also encodes the shape of the entire 100 to 3,800 Hz spectrum since this is useful in the baseband coding. In the first step, the spectrum is divided into four regions illustrated in Fig. 7:

125 - 583 Hz
625 - 1959 Hz
2000 - 3416 Hz
3468 - 3833 Hz

These regions correspond roughly to the usual locations of the first four formants. The dynamic range of the magnitudes of the spectral coefficients is much smaller within each of these regions than in the spectrum as a whole. For voiced phonemes the peak magnitude near 250 Hz can be 30 dB above the magnitudes near 3,800 Hz. The first step of spectral normalization is performed by finding the peak magnitudes within each region, quantizing these peaks to 5 bits each with a logarithmic quantizer, and dividing each spectral coefficient by the quantized peak in its region. The result is a vector of spectral coefficients with maximum magnitude equal to unity. The division into regions should result in the spectral

coefficients being reasonably uniformly distributed within the complex disc of radius one.

The second step extracts more detailed structure. The spectrum is divided into equal bands of about 165 Hz each. The peak magnitude within each band is located and quantized to 3 bits. The complex spectral coefficients within the band are divided by the quantized magnitude and coded to 6 bits each using a hexagonal quantizer. This coding preserves phase information that is important for reconstruction of frame boundaries.

The specifics of this alternative approach are illustrated with reference to Figs. 7 through 11. In this system, the preprocessor 26 is a single-pole pre-emphasis filter. Low frequencies are attenuated by about 5 dB. High frequencies are boosted. The highest frequency (4 kHz) is boosted by about 24 dB. The filter is useful in equalizing the spectrum by reducing the low-pass effects of the initializing filter and the high-frequency attenuation of the lips. The boosting helps to maintain numerical accuracy in the subsequent computation of the Fourier transform.

Within each of the four formant regions, the spectrum is normalized to a curve which in this case is selected as a horizontal line through the peak magnitude of the spectrum in each region. These curves are shown as lines 58, 60, 62 and 64 in Figure 7. The peak magnitude of the complex numbers in each region is determined and encoded to five bits at unit 66 of Fig. 11 by finding a value k which is encoded such that the peak magnitude is between $162 \times 2^{12(k-1)/32}$ and $162 \times 2^{12k/32}$. This results in logarithmic encoding of the peak magnitude. The four k values, each encoded in five bits, make up a total of 20 bits from the formant encoder which are the most significant bits of the transmitted code for the window. All spectral coefficients in each of the four regions are then divided by the $162 \times 2^{12k/32}$ in the spectral

normalization unit 68. By this method, all of the resultant magnitudes, illustrated in Figure 8, are less than 1.

Next, the normalized coefficients output from unit 68 are grouped into 27 regions of four and two subregions of five illustrated in Figure 8. The peak magnitude in each of these subregions is determined and encoded to three bits with a logarithmic quantizer in unit 70. The peak is always coded to the next largest value. The three bits from each of the 22 subregions provide an additional 66 bits of the final signal for the window. Each output within a subregion is multipled by the reciprocal of the quantized magnitude in the sample normalization unit 72, thus ensuring that all outputs illustrated in Fig. 9 remain less than 1.

Each complex output from the baseband of 125 Hz to 1959 Hz of the normalized spectrum of Fig. 9 is coded to six bits with the two dimensional quantizer and encoder 74. The two-dimensional quantizer is formed by dividing a complex disc of radius one into hexagons as shown in Figure 10. The x, y coordinates are radially warped by an exponential function to approximate a logarithmic coding of the magnitude. All points within a hexagon are quantized to the coordinates of the center of the hexagon. As a result, coefficients of large magnitude are coded to better phase resolution than coefficients of small magnitude. Actual quantization is done by table lookup, but efficient computational algorithms are possible.

The bit allocation for a single frame may be summarized as follows:

| | | |
|---|---|---|
| Formant region scale factors | 4 x 5 bits each = | 20 bits |
| Subband scale factors | 22 x 3 bits each = | 66 bits |
| Baseband components | 45 x 6 bits each = | 270 bits |
| TOTAL | | 356 bits |

In a practical 16-kb/s transmission system, this allows 4 bits per frame for overhead functions, such as frame synchronization. The actual coding transformations, bit allocations, and subband sizes may be changed as the coder is optimized for different applications.

All normalization factors (four at 5 bits each, 23 at 3 bits each) and the coded normalized baseband coefficients (45 at 6 bits) are transmitted. At the receiver the baseband is decoded and duplicated into the upper frequency ranged. The normalization factors are applied onto the spectrum to restore the original shape. Specifically, in the receiver, the inverse Fourier Transform Inputs 0 to 2 and 93 to 96 are set to zero. The normalized complex coefficients for Inputs 3 to 47 are reconstructed from the quantizer codes by table lookup. They are duplicated into Positions 48 to 92. This duplication is the nonlinear regeneration step. The scale factors for the subregions and larger regions are then applied.

The inverse transform is computed in unit 36. The effects of the windowing are removed by adding the last 12 points of the previous inverse transform to the first 12 points from the current inverse transform. The speech now passes through filter 38, which is an inverse to the pre-emphasis filter and which attenuates the high frequencies, removing the effects of the treble boost and reducing high-frequency quantization noise. The outputs are converted to analog with a 12-bit linear analog to digital converter 40.

The baseband which is repeated in the spectrum reconstruction has been described as being a band of lower frequencies. However, the baseband may include any range of frequencies within the spectrum. For some sounds where higher energy levels are found in the higher frequencies, a baseband of the higher frequencies is preferred.

It should be noted that the baseband suffers degradations only from quantization errors. The reconstruction of the upper frequencies is only as good as the model and the shaping information. However, by ensuring that at least some coefficient in each 165-Hz band of the normalized baseband is at full scale, each formant is excited at approximately the right frequency. This is an improvement over baseband residual excitation in which some parts of the spectrum may have too little energy. The reduction in computational complexity due to peak finding and scaling instead of linear prediction analysis and filtering is very significant.

This approach is a wideband approach in that the entire voice frequency range is coded. The major problem with other wideband systems at 16 kb/s is that there are barely enough bits available to give a rough description of the waveform. Baseband excitation systems such as the present system meet that problem by devoting most of the bits to the baseband and regenerating the excitation signal for higher frequencies. In a modification of the subband transform coding just described, one could code the baseband as described above, but code only some measure of energy for the higher frequencies. Frequency translation of the baseband regenerates the fine structure of the upper spectrum.

While the invention has been particularly shown and described with reference to a preferred embodiment thereof, it will be understood by those skilled in the art that various changes in form and details may be made

therein without departing from the spirit and scope of the invention as defined by the appended claims.

CLAIMS:

1. A speech encoder comprising:

transform means for performing a discrete transform of an incoming speech signal to generate a discrete transform spectrum of coefficients;

normalizing means for modifying the transform spectrum to provide a normalized, flatter spectrum and for encoding a function by which the discrete spectrum is modified; and

means for encoding at least a portion of the spectrum.

2. A speech coding system as claimed in Claim 1 wherein the normalizing means comprises means for defining the approximate envelope of the discrete spectrum, for digitally encoding the defined envelope and for defining the discrete spectrum relative to the defined envelope to provide a normalized spectrum.

3. A speech coding system as claimed in Claim 2 wherein:

the normalizing means comprises means for defining the approximate envelope of the discrete spectrum in each of a plurality of subbands of coefficients and for encoding the defined envelope of each subband of coefficients and means for scaling each spectrum coefficient relative to the defined envelope of the respective subband of coefficients; and

the means for encoding encodes the scaled spectrum coefficients within each subband in a number of bits determined by the defined envelope of the subband.

4. A speech coding system as claimed in Claim 3 wherein the number of bits determined for a plurality of subbands is zero such that the scaled coefficients for those subbands are not transmitted.

5. A speech coding system as claimed in Claim 4 wherein the scale coefficients of different subbands are encoded in different numbers of bits other than zero.

6. A speech coding system as claimed in Claim 4 wherein the encoded speech is decoded by replicating subbands of transmitted coefficients as substitutes for subbands of nontransmitted coefficients, the transmitted coefficients being replicated such that the nth subband which is transmitted is replicated as the nth subband which is not transmitted.

7. A speech coding system as claimed in Claim 3 wherein the coefficients of different subbands are encoded in different numbers of bits other than zero.

8. A speech coding system as claimed in Claim 2 wherein:

the normalizing means comprises means for defining the approximate envelope of the discrete spectrum in each of a plurality of subbands of coefficients and for encoding the defined envelope of each subband of coefficients and means for scaling each spectrum coefficient relative to the defined envelope of the respective subband of coefficients; and

the means for encoding encodes the scaled coefficients of less than all of the subbands, the encoded scaled coefficients being those

corresponding to the defined envelopes of greater magnitude, with the scaled coefficients of subbands corresponding to defined envelopes of greatest magnitudes being encoded in more bits than coefficients of subbands corresponding to defined envelopes of lesser magnitudes.

9. A speech coding system as claimed in Claim 18 wherein the encoded speech is decoded by replicating subbands of transmitted coefficients as substitutes for subbands of nontransmitted coefficients, the transmitted coefficients being replicated such that the nth subband which is transmitted is replicated as the nth subband which is not transmitted.

10. A speech coding system as claimed in Claim 18 wherein the transform means performs a discrete Fourier transform.

11. A speech coding system as claimed in Claim 2 wherein the normalizing means comprises:

means for determining the maximum magnitude of the discrete spectrum within each of a plurality of regions of the spectrum; and

means for digitally encoding the maximum magnitude of each region; and

means for scaling each coefficient of the discrete spectrum in each region to the maximum magnitude of each region to provide a first set of normalized coefficients.

12. A speech coding system as claimed in Claim 11 wherein the normalizing means further comprises:

means for determining the maximum magnitude of the first set of normalized in each of a plurality of subregions of the spectrum;

means for digitally encoding the maximum magnitude of each subregion; and

means for scaling each output of the first set of normalized outputs to the maximum magnitude of each subregion to provide a second set of normalized outputs.

13. A speech encoder as claimed in Claim 12 wherein each of the maximum magnitudes is logarithmically encoded.

14. A speech encoder as claimed in Claim 12 wherein the maximum magnitude is determined for each of four regions corresponding to the first four formants.

15. A speech encoder as claimed in Claim 12 wherein only a baseband of the normalized spectrum is encoded.

16. A speech coding system as claimed in Claim 2 wherein the transform means performs a discrete Fourier transform.

17. A method of encoding speech comprising:

performing a discrete transform of a window of speech to generate a discrete transform spectrum;

providing a normalized spectrum by defining at least one curve approximating the magnitude of the discrete spectrum, digitally encoding the defined curve and defining the discrete spectrum relative to the defined curve; and

encoding at least a portion of the normalized spectrum.

18. A method of coding speech as claimed in Claim 17 wherein:

the normlized spectrum is provided by defining the approximate envelope of the discrete spectrum in each of a plurality of subbands of coefficients and digitally encoding the defined envelope of each subband of coefficients and scaling each coefficient relative to the defined magnitude of the respective subband of coefficients; and

the scaled coefficients within each subband are encoded into a number of bits determined by the defined envelope of the subband.

19. The method as claimed in Claim 18 wherein the discrete transform is a Fourier transform.

20. The method as claimed in Claim 19 wherein the number of bits determined for a plurality of subbands is zero such that the scaled coefficients for those subbands are not transmitted.

21. The method as claimed in Claim 20 wherein the scaled coefficients of different subbands are encoded in different numbers of bits other than zero.

22. The method as claimed in Claim 20 wherein the encoded speech is decoded by replicating subbands of transmitted coefficients as substitutes for subbands of nontransmitted coefficients, the transmitted coefficients being replicated such that the nth subband which is transmitted is replicated as the nth subband which is not transmitted.

23. A method as claimed in Claim 17 wherein the normalized spectrum is provided by;

determining a maximum magnitude of the discrete spectrum within each of a plurality of regions of the spectrum;

digitally encoding the maximum magnitude of each region; and

scaling each coefficient of the discrete spectrum in each region to the maximum magnitude of each region to provide a set of normalized coefficients.

24. In a system in which a discrete signal is divided into a plurality of subbands of coefficients and only select subbands of coefficients are transmitted to a receiver as determined by the signal itself, a method of regenerating the discrete signal at the receiver comprising replicating subbands of transmitted coefficients as substitutes for subbands of nontransmitted coefficients, the transmitted coefficients being replicated such that the nth subband which is transmitted is replicated as the nth subband which is not transmitted.

25. A system as claimed in Claim 24 wherein the coefficients are the coefficients of a Fourier transform spectrum of speech.

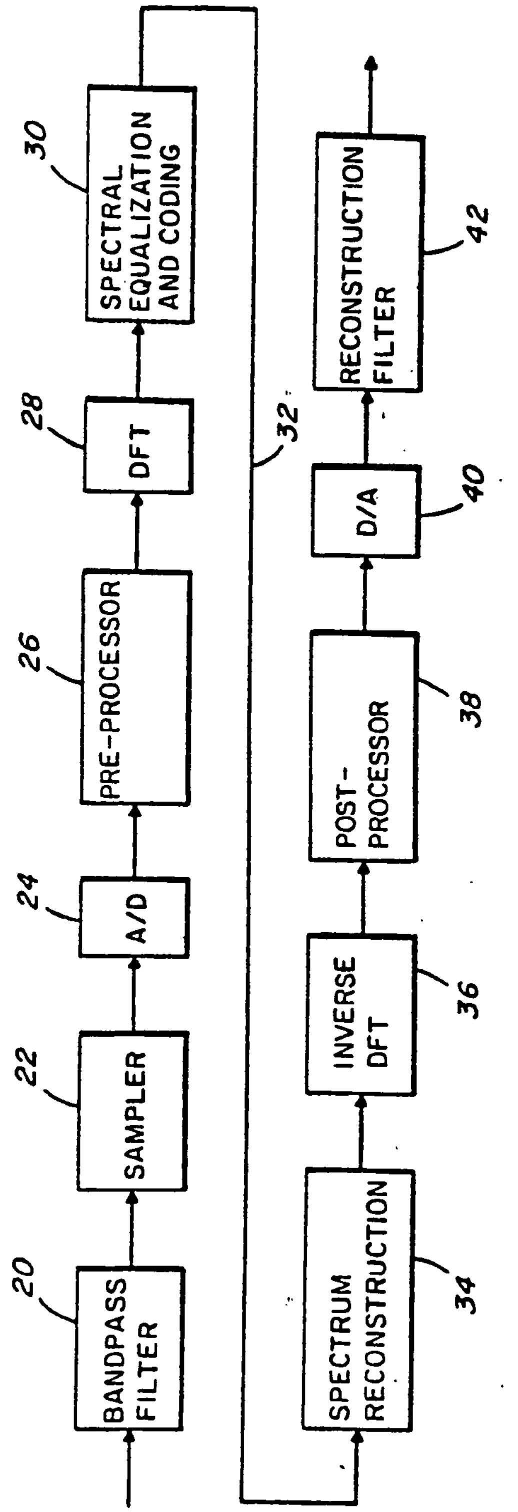

BANDPASS FILTER
20
SAMPLER
22
A/D
24
PRE-PROCESSOR
26
DFT
28
SPECTRAL EQUALIZATION AND CODING
30
32
SPECTRUM RECONSTRUCTION
34
INVERSE DFT
36
POST-PROCESSOR
38
D/A
40
RECONSTRUCTION FILTER
42

FIG. 1

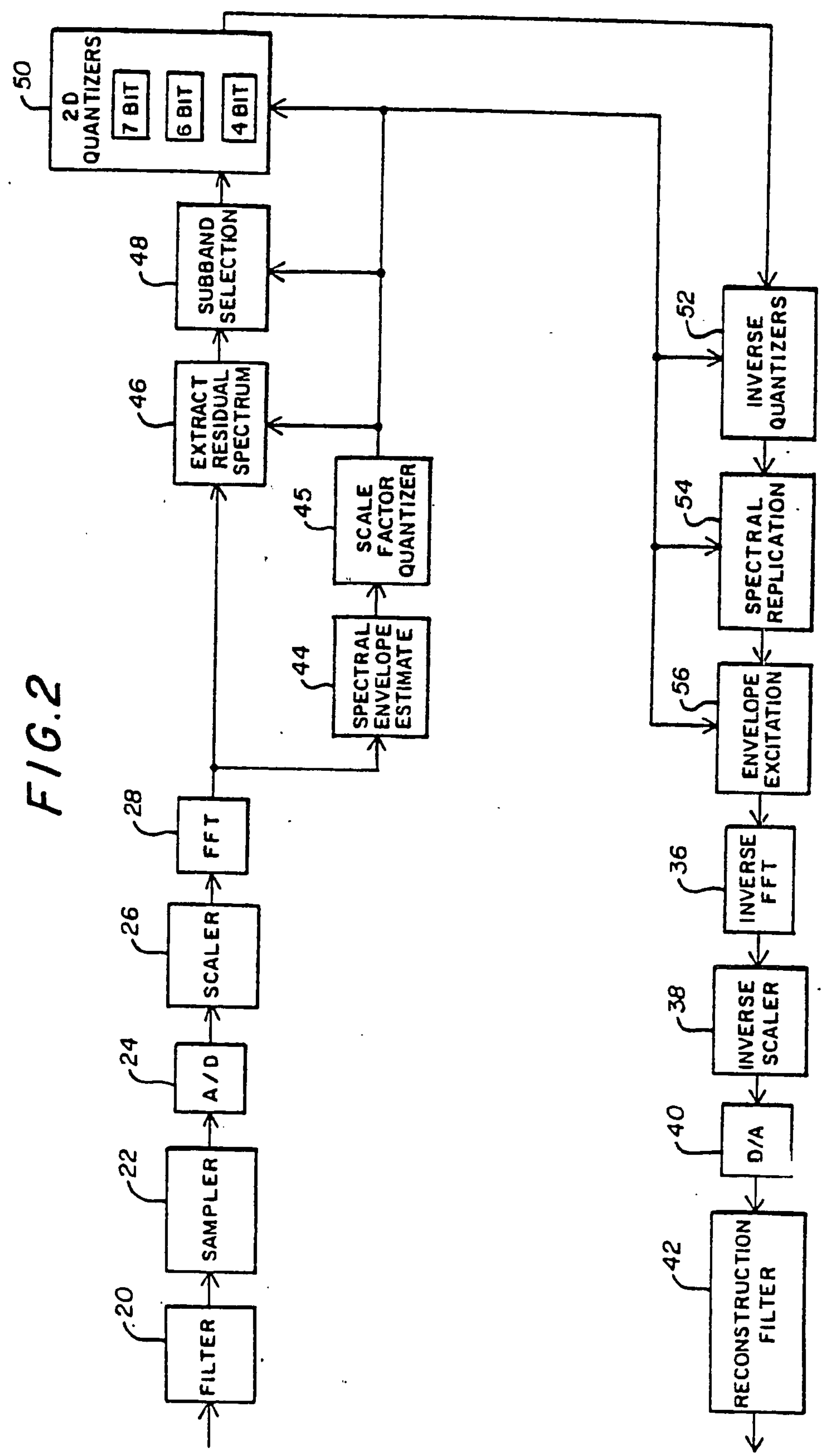

FILTER
20
SAMPLER
22
A/D
24
SCALER
26
FFT
28
SPECTRAL ENVELOPE ESTIMATE
44
SCALE FACTOR QUANTIZER
45
EXTRACT RESIDUAL SPECTRUM
46
SUBBAND SELECTION
48
2D QUANTIZERS
7 BIT
6 BIT
4 BIT
50
INVERSE QUANTIZERS
52
SPECTRAL REPLICATION
54
ENVELOPE EXCITATION
56
INVERSE FFT
36
INVERSE SCALER
38
D/A
40
RECONSTRUCTION FILTER
42

FIG.2

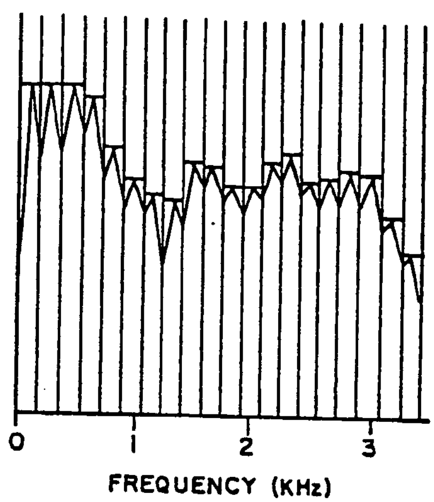



FIG. 3

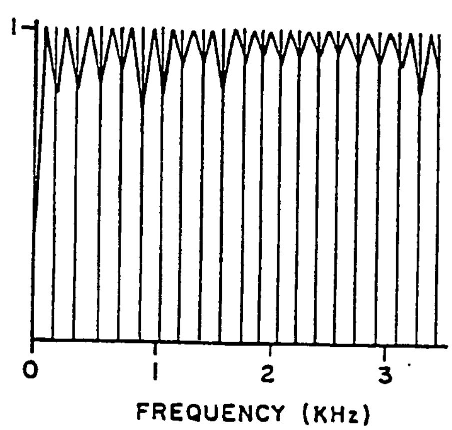
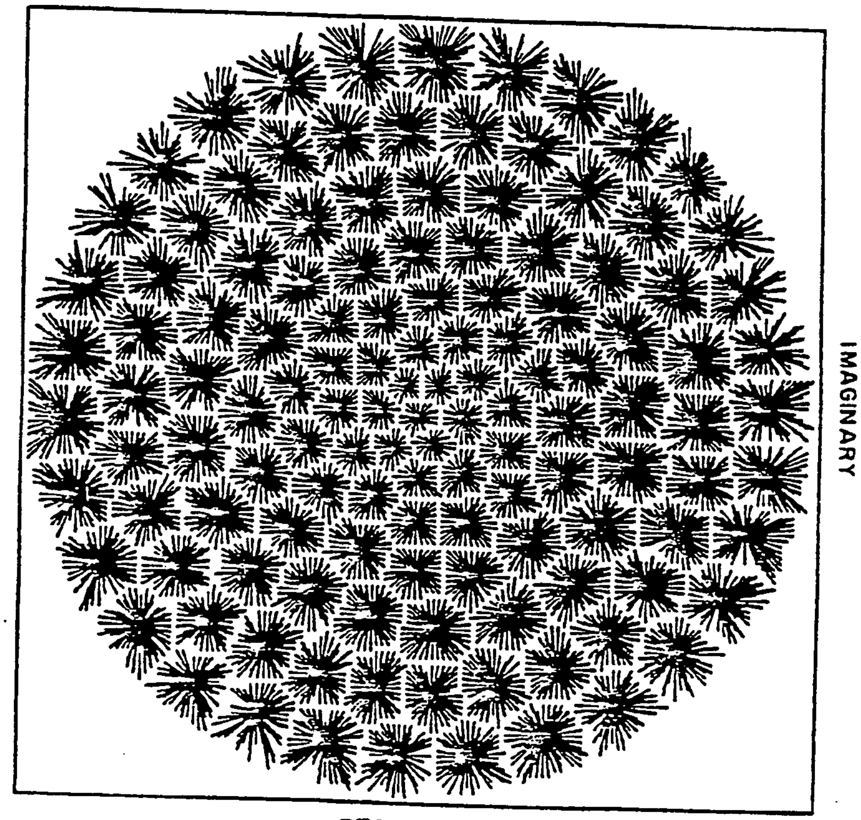



FIG. 4

FIG.5



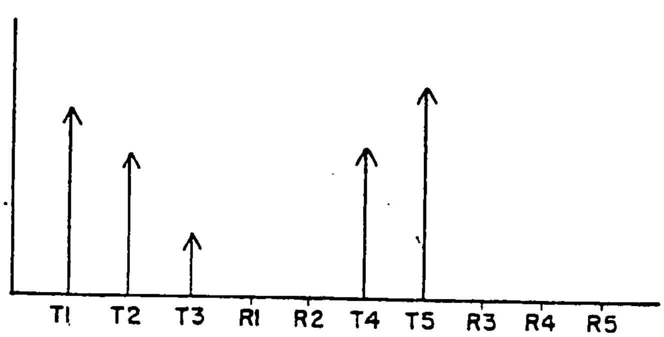




FIG.6

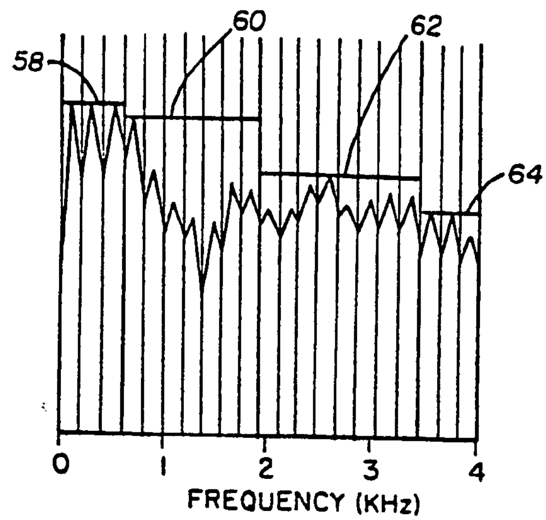



FIG.7

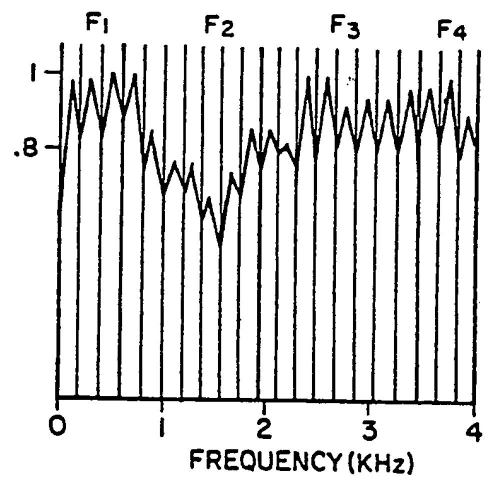



FIG.8

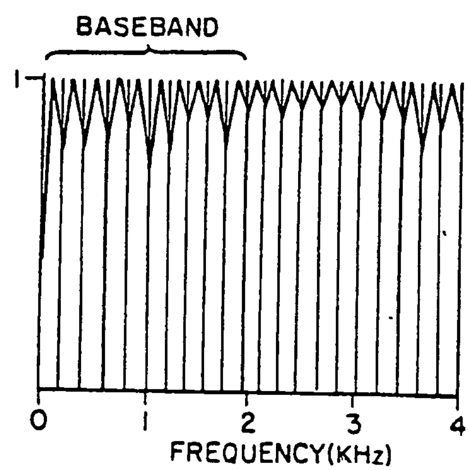



FIG.9

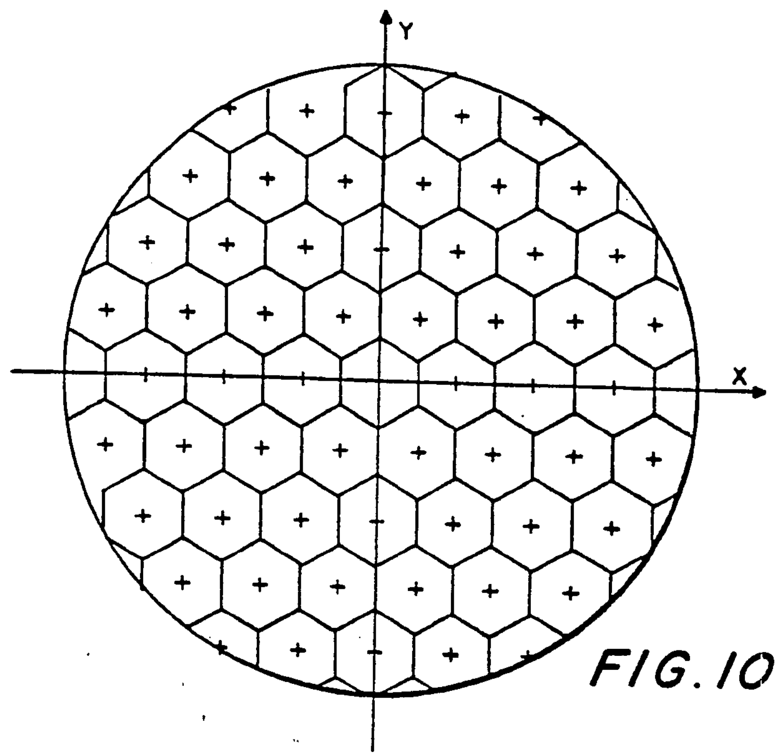

Y
X

FIG. 10

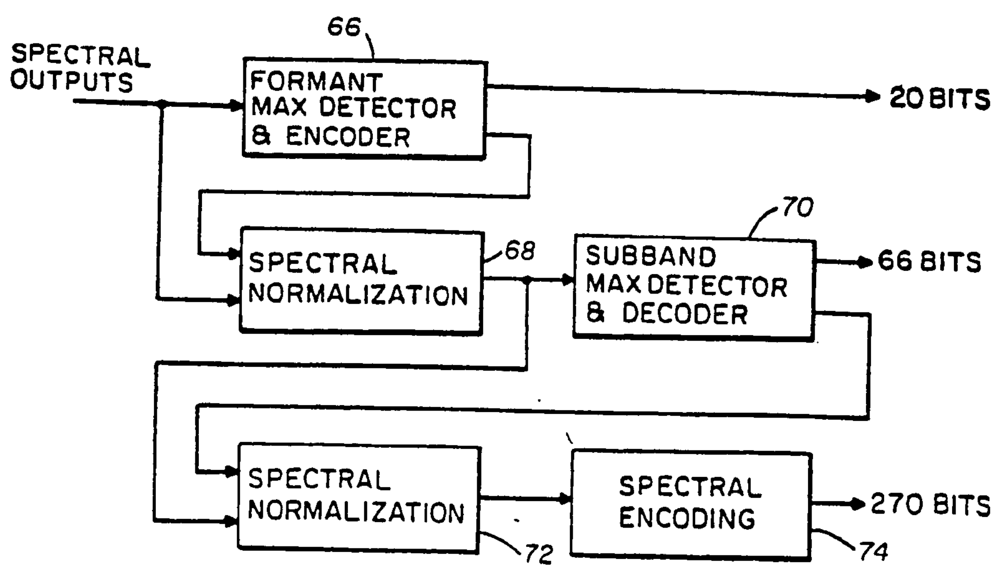

SPECTRAL OUTPUTS
66
FORMANT MAX DETECTOR & ENCODER
20 BITS
SPECTRAL NORMALIZATION
68
70
SUBBAND MAX DETECTOR & DECODER
66 BITS
SPECTRAL NORMALIZATION
SPECTRAL ENCODING
270 BITS
72
74

FIG. 11

# INTERNATIONAL SEARCH REPORT

International Application No PCT/US85/02448

**I. CLASSIFICATI N OF SUBJECT MATTER** (if several classification symbols apply, indicate all) [3]

According to International Patent Classification (IPC) or to both National Classification and IPC
U.S.: 381-31
INT. CL.4 G10L 5/00

**II. FIELDS SEARCHED**

Minimum Documentation Searched [4]

| Classification System | Classification Symbols |
|---|---|
| U.S. | 381-29, 30, 31 |

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [5]

**III. DOCUMENTS CONSIDERED TO BE RELEVANT** [14]

| Category * | Citation of Document, [16] with indication, where appropriate, of the relevant passages [17] | Relevant to Claim No. [18] |
|---|---|---|
| Y | US, A, 4,330,689, Kang, 18 May 1982 | 1-24 |
| Y | IEEE International Conference on Acoustics, 1981, Kang et al, "Mediumband Speech Processor", see pages 820-823. | 1-24 |
| Y,P | US, A, 4,535,472, Tomcik, 13 August 1985 | 1-8, 16-23 |
| Y | US, A, 4,388,491, Ohta et al, 14 June 1983 | 1-3 |

* Special categories of cited documents: [15]
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

**IV. CERTIFICATION**

| Date of the Actual Completion of the International Search [2] | Date of Mailing of this International Search Report [2] |
|---|---|
| 4 February 1986 | 07 MAR 1986 |
| **International Searching Authority** [1] | **Signature of Authorized Officer** [20] |
| ISA/US | Martin Yuen |

Form PCT/ISA/210 (second sheet) (October 1981)